

Что такое статистика?

Немного истории

*«Исчислите всё общество сынов Израилевых по родам их, по семействам их, по числу имен, всех мужского пола поголовно...»
Четвертая книга Моисеева.
Числа. Ветхий завет, гл. 1,2.*

У истоков статистической науки стояли две школы – **немецкая описательная** и **английская школа политических арифметиков**.

Представители описательной школы считали, что задачей статистики является описание достопримечательностей государства: территории, населения, климата, вероисповедания, ведения хозяйства и т. п. только в словесной форме, без цифр и вне динамики, т. е. без отражения особенностей развития государств в различные периоды, а только лишь на момент наблюдения.

Видными представителями описательной школы были, например, **Г. Конринг** (1606–1661), **Г. Ахенваль** (1719–1772), **А. Бюшинг** (1724–1793).

Политические арифметики ставили целью изучать общественные явления с помощью числовых характеристик. От описания явлений и процессов статистика перешла к их измерению, исследованию, оценке и выработке вероятных путей будущего развития. Политические арифметики видели основное назначение статистики в изучении массовых общественных явлений. Виднейшим представителем и основателем этого направления был **В. Петти** (1623–1687). Именно школа политических арифметиков стала основообразующей в развитии современной статистики.

Математическое направление в статистике развивалось в работах англичан – сэра **Фрэнсиса Гальтона** (1822–1911), **Карла Пирсона** (1857–1936), **Рональда Фишера** (1890–1962).

Прогрессу статистической методологии способствовали труды российских статистиков – **А.А. Чупрова** (1874–1926), **В.С. Немчинова** (1894–1964), **С.Г. Струмилина** (1877–1974), **В.Н. Старовского** (1905–1975) и др.

Много определений

«Statistics is the logic of uncertainty».
Probability lecture (Harvard)

В настоящее время термин «статистика» употребляется в основном в трех значениях.

Во-первых, под статистикой понимают **отрасль практической деятельности**, которая имеет своей целью **сбор, обработку, анализ и публикацию массовых данных о самых различных явлениях общественной жизни**. Осуществляется сбор данных в каждом регионе и по стране в целом о численности и составе населения, ведется подсчет предприятий и организаций, собираются данные об объемах производства и объемах продаж и т.д. Эту деятельность на профессиональном уровне осуществляет **Федеральная служба государственной статистики (Госкомстат РФ)** и система ее учреждений, организованных по административно-территориальному признаку.

Во-вторых, статистикой называют **цифровые материалы**, служащие для характеристики какой-либо области общественных явлений или территориального распределения какого-то показателя, публикуемые в периодической прессе, справочниках, сборниках. Например, динамика цены на бензин в Томской области, представленная за второе полугодие 2008 года.

В-третьих, статистикой называется **отрасль знания, особая научная дисциплина**, которая в широком понимании **разрабатывает методы сбора, систематизации, анализа, интерпретации и отображения результатов наблюдений массовых случайных явлений и процессов с целью выявления существующих в них закономерностей**. Например, исследования взаимосвязи между качеством трудовых ресурсов и экономическим ростом в регионах РФ.

Приведем несколько определений статистики.

«С моей точки зрения, собственно статистику можно, пожалуй, определить как искусство оперирования с неопределенностью и различиями в поведении людей в ситуациях принятия решений».

(L.J. Savage. The Foundations of Statistics, 1954).

«Статистика как наука является одним из разделов прикладной математики, и ее можно рассматривать как математику, применяемую при разработке результатов массового наблюдения... Статистику можно рассматривать как: 1) учение о совокупностях, 2) учение о вариации и 3) учение о методах приведения данных в компактной форме».

(Р.А. Фишер. Статистические методы для исследователей, 1958).

«В наше время принято считать, что статистика есть наука, изучающая теорию принятия решений в условиях неопределенности. Это определение статистики выкристаллизовывалось в результате многих лет ее развития. Достоинство этого определения состоит в том, что оно в сжатой и ясной форме излагает научное существо статистики».

(Г. Чернов, Л. Мозес. Элементарная теория статистических решений, 1962).

«Статистику иногда определяют как искусство и науку количественной обработки наблюдений, подверженных изменениям».

(E.V. Lewis. Statistical Analysis. Ideas and Methods, 1963).

«Значение этого слова (статистика) за последние два столетия претерпело значительные изменения. Слово "статистика" имеет один корень со словом "государство" (state) и первоначально оно означало искусство и науку управления; первые преподаватели статистики университетов Германии XVIII в. сегодня назывались бы специалистами по общественным (политическим) наукам. Поскольку решения правительства до некоторой степени основываются на данных о населении, промышленности (ремеслах), сельском хозяйстве и т. д., статистики, естественно, стали интересоваться такими данными, и постепенно слово "статистика" стало означать сбор данных о государстве, а затем вообще сбор и обработку данных. В этом значении слово все еще широко употребляется, но наблюдается и дальнейшее изменение значения. Нет смысла собирать данные, если из этого не извлекается какая-то польза, и статистики, естественно, начинают заниматься интерпретацией данных. Современный статистик изучает методы, при помощи которых можно сделать выводы о популяции на основе данных, которые обычно получают из выборки популяции».

(J.L. Hodges, E.L. Lehmann. Basic Concepts of Probability and Statistics, 1964).

«В основе статистического вывода лежит индуктивное рассуждение, приводящее к утверждениям, верным лишь с определенной "степенью достоверности". Строгий смысл "степени достоверности" обеспечивается математическими методами и вероятностным подходом, образующими основы современной статистической теории».

(С.Р. Рао. Линейные статистические методы и их применение, 1968).

«Математическая статистика – это ветвь теории вероятностей. В ней рассматриваются задачи, связанные с оперативными характеристиками правил индуктивного поведения, основанных на случайных экспериментах».

(Ю. Нейман. Вводный курс теории вероятностей и математической статистики, 1968).

«Статистика – это математическая теория того, как узнать нечто о мире через опыт».

(W. Thompson. The Future of Statistics, 1968).

В современных учебниках по теории статистики часто приводится следующее определение.

«Статистика – это научно упорядоченное изображение действительности, вызванное необходимостью совершенствовать искусство управления».

Перечислим основные этапы статистического анализа.

1. Планирование исследования включает составление подробного плана сбора данных, возможно, с использованием случайной выборки из генеральной совокупности.

2. Предварительное исследование данных включает рассмотрение набора данных с разных точек зрения, описание и обобщение данных. Выполнение этого этапа помогает убедиться, что запланированный анализ адекватен данным, а при необходимости позволяет внести в процесс анализа определенные коррективы.

3. Оценивание неизвестной величины дает наиболее обоснованное возможное предположение о значении, основанное на исходных данных. Кроме того, есть возможность вычислить величину ошибки, которая возникает при использовании оценки вместо фактического неизвестного значения.

4. Проверка статистических гипотез заключается в использовании данных для выбора одной из двух (или больше) различных возможностей при решении вопроса в неопределенной ситуации. Такая проверка позволяет убедиться, действительно ли данные обладают определенным интересным свойством, или мы имеем дело с «чистой случайностью», которая не представляет интереса.

Все статистические выводы опираются на понятие вероятности.

Вероятность, исходя из предположений об изучаемой ситуации, показывает возможность или шанс наступления в будущем каждого из нескольких потенциальных событий. **Вероятность** – это понятие, в некотором смысле обратное статистике: вероятность показывает, какие данные вы скорее всего получите, если известна характеристика ситуации, а статистика помогает охарактеризовать ситуацию в результате анализа и обобщения данных.

Исходным материалом для статистики являются данные (результаты наблюдений за интересующем нас явлением).

Набор **данных** содержит одно или несколько значений для каждого из отдельных объектов, называемых **элементарными единицами**. В качестве таких объектов могут выступать люди, домохозяйства, города, телевизионные приемники или что угодно, что представляет интерес для изучения. Для каждого из объектов регистрируют один и тот же признак (или признаки). Признак, который регистрируют для каждого из объектов (например, стоимость), называется переменной.

Существуют **три основных способа классификации наборов данных**: по количеству переменных (одномерный, двумерный и многомерный); по типу представленной каждой из переменных **информации** (числа или категории) и в зависимости от того, является ли набор данных **временным рядом**, или это **данные об одном временном срезе**.

Одномерные наборы данных (одна переменная) содержат информацию только об одном признаке, зарегистрированную для каждого объекта. Одномерный набор данных позволяет определить типичное значение и характеристику изменчивости данных, а также выделить специфические особенности или проблемы в данных.

Двумерные наборы данных (две переменные) содержат два признака, значения которых регистрируются для каждого объекта. Двумерные данные в дополнение к информации о каждой переменной как наборе одномерных данных позволяют изучить **связь между двумя переменными** и предсказать значение одной переменной на основе значения другой.

Многомерные наборы данных (много переменных) содержат три или больше признаков, значения которых регистрируются для каждого объекта. Многомерные данные в дополнение к информации о каждой переменной как наборе одномерных данных дают возможность изучить связь между переменными и предсказать значение одной переменной на основе значения других.

Значения переменных, которые регистрируются как числа, имеющие содержательный смысл, называют **количественными данными**. Дискретная количественная переменная может принимать значения только из некоторого списка чисел (таких как, например, 0 или 1, или перечень чисел 0, 1, 2, 3, ...). Любую количественную переменную, которая не является **дискретной**, будем называть **непрерывной**. Значения непрерывной переменной невозможно задать конечным или счетным списком значений.

Если переменная содержит информацию о том, какой из нескольких нечисловых категорий принадлежит объект, то она называется **качественной переменной**. Если категории можно естественным образом и осмысленно упорядочить, то речь идет о **порядковой качественной переменной**. Если такой порядок отсутствует, то речь идет о **номинальной качественной переменной**. Несмотря на то, что часто значения качественной переменной можно записать (закодировать) с помощью чисел, такая переменная все равно остается качественной, а не количественной, поскольку эти числа не имеют какой-либо интерпретации, содержательно присущей этой переменной.

К количественным данным можно применять те же операции, что и к обычным числам: подсчет частоты, ранжирование, арифметические действия. С порядковыми данными можно выполнять только подсчет частоты и ранжирование, с номинальными данными – только подсчет частоты.

Если последовательность записи данных имеет содержательный смысл, то соответствующий набор данных представляет собой **временной ряд**. Если последовательность записи данных не важна, то соответствующий набор содержит **данные об одном временном срезе**. Анализ временных рядов сложнее анализа данных об одном временном срезе.

Если вы самостоятельно планируете сбор данных, то получаете **первичные данные**. Если вы используете данные, предварительно собранные другими людьми и для других целей, то вы имеете дело с **вторичными данными**. Получение первичных данных часто обходится дорого и занимает много времени, но вы получаете то, что вам необходимо. Вторичные данные можно получить дешевле (или даже бесплатно), но вы можете найти, а можете и не найти то, что вам необходимо.

Зачем статистика менеджеру?

«В жизни, как правило, преуспевает тот, кто располагает лучшей информацией».

Бенджамин Дизраэли

В менеджменте постоянно используют данные. Ниже приведен краткий перечень видов ежедневно используемой менеджерами информации.

1. Финансовые отчеты (и другие виды бухгалтерской отчетности).
2. Курсы и объемы ценных бумаг, процентные ставки (и другая информация, относящаяся к инвестициям).
3. Состояние бюджета (и другие сообщения правительства).
4. Отчеты о продажах (и другие внутренние отчеты).
5. Результаты обзора состояния рынка (и другие маркетинговые отчеты).
6. Данные о качестве продукции (и другие производственные отчеты).
7. Отчеты о производительности рабочих (и другие внутренние данные фирмы).
8. Цена и объем проданной продукции (и другие данные о продажах).
9. Расходы на рекламу и результаты рекламной компании (и другая рекламная информация).

Возможно, кто-то будет консультировать вас по этим вопросам. Вы редко будете видеть фактические данные. Время от времени вы могли бы попросить дать вам «сырые данные». Просмотрев их, можно получить неожиданные результаты: возможно качество данных не так высоко, как вы думали (и у вас появится мысль: на чем же мы строим свои прогнозы?), или, напротив, вы обретете уверенность. Другими словами, **данные заслуживают внимания!**

Статистические результаты должны допускать простое непосредственное объяснение (даже если соответствующая теория намного сложнее). Дадим несколько рекомендаций.

1. Доверяйте своим суждениям, учитывайте здравый смысл.
2. Сохраняйте здоровый скептицизм.
3. Не дайте себя ввести в заблуждение с помощью на первый взгляд оригинального статистического анализа. Он может опираться

на нереальные или неподходящие предположения.

Вообще говоря, из-за большой гибкости, доступной аналитику на каждой стадии статистического анализа, один из самых важных факторов, который надо принять во внимание при оценке результатов статистического анализа, звучит так: Кто это финансировал? Помните, что аналитик много раз делает выбор – при определении проблемы, при планировании сбора данных, при выборе структуры или модели для анализа, при интерпретации результатов.

Как пишет Мхитарян Владимир Сергеевич, заведующий кафедрой статистических методов ВШЭ, профессор, «...с каждым годом в нашей стране и мире растет потребность в экономистах, обладающих аналитическим мышлением, основанном на глубоком знании экономики, математики, статистики и компьютерных технологий. Значительная потребность в статистиках-аналитиках имеется на сегодня на микроэкономическом уровне у государственных и частных предприятий, учреждений и фирм. В XXI веке, по мнению американских экспертов, системный аналитик (статистик) входит в четверку самых востребованных специальностей в мире наряду с финансовым менеджером, менеджером веб-сайтов и специалистом по информационным технологиям. Ведущие университеты мира, занимающие первые места в мировых и национальных рейтингах (Гарвард, Стенфорд, Оксфорд, Кембридж и др.), имеют в своей структуре факультеты (кафедры, отделения, школы) статистики.

Уже сегодня многие министерства в России, а также крупные отечественные и зарубежные компании имеют аналитические подразделения, занимающиеся анализом статистических данных для определения направлений и перспектив своего развития».

Основные термины

- **Статистика (statistics)**
- **Планирование исследования (designing the study)**
- **Предварительное исследование данных (exploring the data)**
- **Оценка неизвестной величины (estimating an unknown quantity)**
- **Проверка статистических гипотез (hypothesis testing)**
- **Вероятность (probability)**
- **Набор данных (data set)**
- **Элементарные единицы (elementary units)**
- **Переменная (variable)**
- **Одномерный (univariate)**
- **Двумерный (bivariate)**

- **Многомерный (multivariate)**
- **Количественная (quantitative)**
- **Дискретная (discrete)**
- **Непрерывная (continuous)**
- **Качественная (qualitative)**
- **Порядковая или ординальная (ordinal)**
- **Номинальная (nominal)**
- **Временные ряды (time series)**
- **Об одном временном срезе (cross-sectional)**
- **Первичные данные (primary data)**
- **Вторичные данные (secondary data)**